Open Elective for BE (E&TC/Electronics)

404190: Data Science & Analytics

Teaching Scheme:
Lectures: 4 Hrs. / Week
Tutorial: 1 Hr. /Week

Examination Scheme:
In Semester Assessment: 30
End Semester Examination: 70

Course Overview:

This course provides foundation level training that enables immediate and effective participation in big data and other analytics projects. It includes an introduction to database management and big data analytics. The course provides grounding in basic analytic methods and an introduction to big data analytics technology and tools.

Course Objectives:
Upon successful completion of this course, participants should be able :
- To understand the fundamental concepts of database management. These concepts include aspects of database design, database languages, and database-system implementation.
- To learn a powerful, flexible and scalable general purpose database to handle big data.
- To Deploy the Data Analytics Lifecycle to address big data analytics projects
- To Apply appropriate analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results
- To Select appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences

Unit I: Data Base Management System   (8 Hrs.)
Introduction to Database Management Systems, Purpose of Database Systems, Database-System Applications, View of Data, Database Languages, Database System Structure. Introduction to NoSQL Database, Types and examples of NoSQL Database- Key value store, document store, graph, Performance, Structured verses unstructured data, Comparative study of SQL and NoSQL.

Unit II: SQL AND PL/SQL   (8 Hrs.)
Characteristics and advantages, SQL Data Types and Literals, DDL, DML, DCL, TCL, SQL Operators, Tables: Creating, Modifying, Deleting, Views: Creating, Dropping, Updating using Views, Indexes, SQL DML Queries. **PL/SQL:** Concept of Stored Procedures & Functions, Cursors, Triggers, Assertions, roles and privileges , Embedded SQL, Dynamic SQL.

Unit III: Introduction to Big Data Science   (6 Hrs.)
Definition of Big Data, Big data characteristics & considerations, Data repositories- analyst perspective, Business drivers for analytics, Typical analytical architecture, Business Intelligence Vs Data science, Drivers of Big data analytics, Role of data scientist in Big data ecosystem, Applications of Big data analytics.

Unit IV: Data Analytics Lifecycle   (6 Hrs.)
Need of Data analytic lifecycle, Key roles for successful analytic projects, **various phases of Data analytic lifecycle:** Discovery, Data Preparation, Model Planning, Model Building, Communicating Results, Operationalization.

Unit V: Basic Data Analytics methods using R   (8 Hrs.)
Introduction to R: GUI of R, Getting data into & out of R, Data types in R, Basic operations, Basic statistics, Generic functions, Data visualization using R, Data exploration & presentation, Statistics for model building & evaluation.

Unit VI: Advanced Analytics- Theory & Methods   (8 Hrs.)
K-means Clustering, Association Rules, Apriori algorithm, Linear Regression, Logistics Regression, Naïve Bayesian classifiers, Decision Trees, Time series analysis, Text analysis.

Text Books
1. Silberschatz A., Korth H., Sudarshan S., "Database System Concepts", McGraw Hill Publishers, ISBN 0-07-120413-X, 6th edition
2. David Dietrich, Barry Hiller, "Data Science & Big Data Analytics", EMC education services, Wiley publications, 2012
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning", Springer, Second Edition, 2011.

Reference Books
1. C J Date, "An Introduction to Database Systems", Addison-Wesley, ISBN: 0201144719
2. Mark gardner, "Beginning R: The Statistical Programming Language", Wrox Publication
3. Adam Fowler, "NoSQL For Dummies", John Wiley & Sons, ISBN-1118905628

Guidelines for Tutorial

Tutorials must be conducted batch wise. Batch size should not be more than 20 students. The main objective of this tutorial is to focus on the outcomes defined in the theory syllabus by solving the following assignments based on paper work.

List of Tutorials:

Perform any 8 Tutorials from following

1. Create a Table with single Pivot Key Id and three other fields with data type varchar.
2. Create a procedure to update a table using where claus.
3. Perform Basic operation on R.
4. Plotting of a table Using R.
5. Conversion of unstructured data into structured data.
6. Cleansing of data from a Table using R.
7. Apply linear regression on structured data.
8. Apply k-means clustering on structured data.
9. Perform Sentimental Analysis Using R.
10. Perform Time Series analysis Using R.